

**Project Report**

**Real Estate Price Prediction(Boston)**

**Course Title:**

Statistical Data Analysis (DS Major)

**Course Code:** DS-411

**Submitted to**

Ms. Maksuda Akter Rubi

Department of Software Engineering

Daffodil International University

**Submitted By**

Name: MD. Rakibuzzaman

Id: 181-35-347

Sec: PC-A

Department of Software Engineering

Daffodil International University

**Contents**

|  |  |  |
| --- | --- | --- |
| *SR. NO.* | *CHAPTER NAME* | *PAGE NO.* |
| 01. | Abstract | 03 |
| 02. | Introduction | 03 |
| 03. | Exploratory Data Analysis | 04 |
| 04. | Methodology | 06 |
| 05. | Conclusion | 07 |
| 06. | Reference | 07 |

**Abstract**

The real estate market is a standout amongst the most focused regarding pricing and keeps fluctuating. It is one of the prime fields to apply the ideas of machine learning on how to enhance and foresee the costs with high accuracy. There are three factors that influence the price of a house which includes physical conditions, concepts and location. The current framework includes estimating the median value of owner-occupied homes in $1000s. The objective of the paper is prediction of residential prices for the customers considering their financial plans and needs. By breaking down past market patterns and value ranges, and coming advancements future costs will be anticipated. This examination means to predict house prices median value with Multiple Linear Regression or Neural Network Model. The Multiple Linear Regression shows us the adjusted R squared value is 0.5846. It will help clients to put resources into a bequest without moving toward a broker. The result from this research proved Neural Network model gives minimum prediction rmse score which is 0.2436096.

**Introduction**

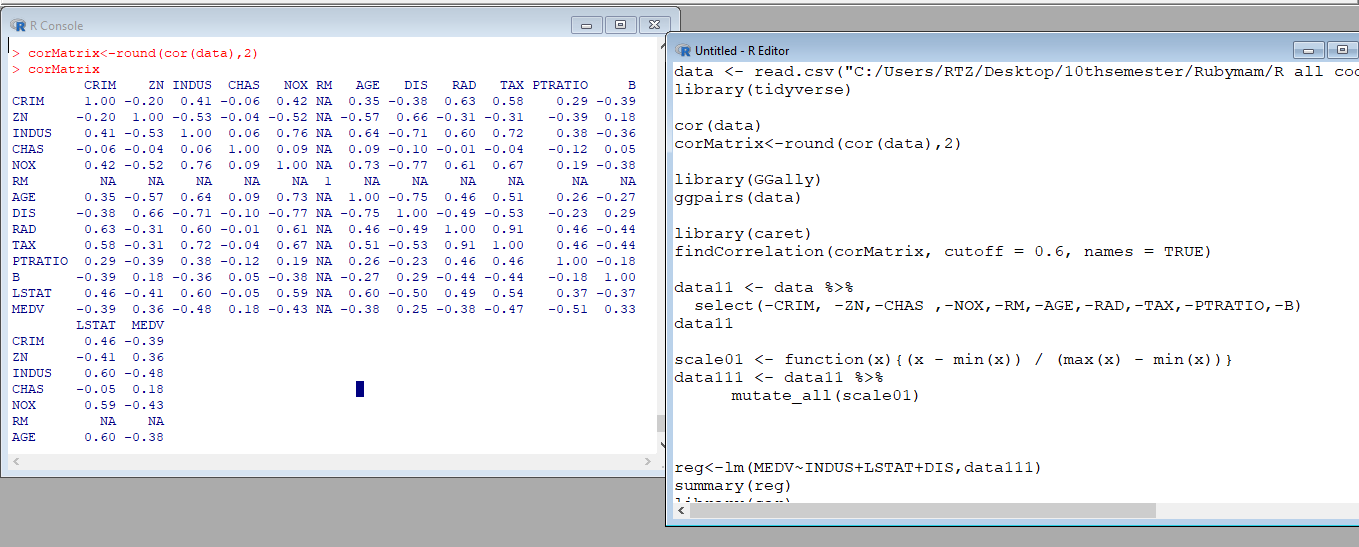
A variety of techniques have been developed for carrying out the valuations, one of the principal approaches being the market comparison essentially establishing the value of a property by comparisons with similar properties. Considered here is the situation of a real estate agent being engaged by a vendor to provide a market valuation of a house. In practical terms this involves prediction of the selling price with different losses being associated with different prediction errors.

Considered here is the situation of a real estate agent being engaged by a vendor to provide a market valuation of a house. In practical terms this involves prediction of the selling price with different losses being associated with different prediction errors. The loss from overestimation would generally be different from that of underestimation by the same amount, and its magnitude typically different from the monetary amount of the error. The consequences of over and underestimation are discussed and an attempt is made to elicit the real estate agent's loss function. It is shown that if the root mean squared error is measure by rmse score using the predictive mean--the adjustment being a fraction of the standard deviation, or variance, of the prediction errors.

**Exploratory Data Analysis**

Exploratory Data Analysis is a very important step before training the model. In this section, I will use some visualizations to understand the relationship of the target variable with other features. I will use R programming for visualize and exploratory my data analysis related things.

I have used cor method to observe the correlation between dependent and independent variables.



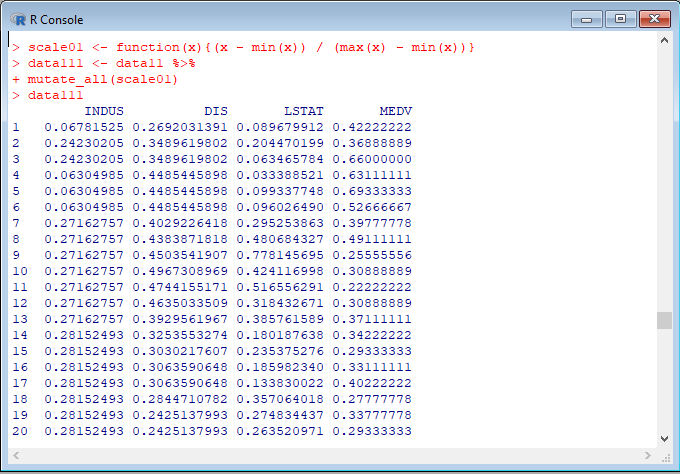
Let’s first plot the distribution of the target variable MEDV. I will use the ggpairs(data) function from the GGally library.



Looking for Correlation:

The correlation coefficient ranges from -1 to 1. If the value is close to 1, it means that there is a strong positive correlation between the two variables. When it is close to -1, the variables have a strong negative correlation. So we can say that the values of this dataset are linearly compatible with each other. I have choose data which is greater than 0.6 or positively correlated with the dependent variable. I removed the variables which have the collinearity problems and was not significant enough. I have used caret library to findout the upper value of 0.6 to cutoff the variables. Then I have scalled my data to get a proper outcome.

Scalled and Cleaning Data:



**Methodology**

The Housing dataset which contains information about different houses in Boston. This data was originally a part of UCI Machine Learning Repository and has been removed now. I can also access this data from the library. There are 506 samples and 13 feature variables in this dataset. The objective is to predict the value of prices of the house using the given features.

The consequences of over and underestimation are discussed and an attempt is made to elicit the real estate agent's loss function. It is shown that if the rmse score by Neural Network Model and measure adjusted R squared value using Multiple Linear Regression and the distribution of prediction errors normal, the expected loss is minimised by additively adjusting the predictive mean--the adjustment being a fraction of the standard deviation, or variance, of the prediction errors. Although the paper deals with real estate, its applicability is much wider.

***Multiple Linear Regression*:**

I have used multiple liner regression model to train it on both train and test sets. The adjusted R squared value is 0.5846.

***Neural Network Model***:

I have also used Neural Network Model to train our model on both the training and test sets. And the rmse score is 0.2436096.

**Conclusion**

In this project I applied the concepts of Neural Network Model on the real estate price prediction dataset. An integrated methodology is presented for determining the optimal prediction of the response variable in a standard linear model. Various types of asymmetry of the loss function are considered. Theoretical developments are presented, deriving adjustments to the predictive mean in order to minimise expected loss in the presence of the asymmetry. Tables of values of adjustments are given, to be used in predicting the response. In the process, an intermediate step is the elicitation of the parameters of both the loss function and prior distribution of an expert decision maker, in this case a real estate agent active in the market. Neural Network Model is considered and used rmse score for measuring the performance. The particular application is to real estate valuation the methodology is applicable to a wide variety of problems of prediction.

**References**

* Broemeling, L. D. (1985). Bayesian Analysis of Linear Models, Marcel Dekker, New York. Cain, M. (1991). Minimal quantile measures of predictive cost with a generalized cost of error function, Economic Research Papers, No. 40, University of Wales, Aberystwyth.
* Cain, M. and Owen, R. J. (1990). Regressor levels for Bayesian predictive response, J. Amer. Statist. Assoc., 85~ 228-231.
* Granger, C. W. J. (1969). Prediction with a generalized cost of error function, Operational Research Quarterly, 20, 199-207.
* Jeffreys, H. (1961). Theory of Probability, Clarendon Press, Oxford, U.K.
* Savage, L. J. (1962). Bayesian Statistics: In Decision and Information Processes, Macmillan, New York.
* Varian, H. R. (1975). A Bayesian approach to real estate assessment, Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage (eds. S. E. Fienberg and A. Zellner), 195 208, North-Holland, Amsterdam.